Security for Collaboration in Open, Scientific Computing Environments

A Summary of
The 4th, Joint DOE
Office of Science - Office of Defense Programs Laboratories
Cybersecurity Workshop
Hyatt Regency O'Hare, Ill., Jan. 17-18, 2001

(http://www.itg.lbl.gov/DOE_Security_Research)

Workshop Participants

William E. Johnston, Convener Lawrence Berkeley National Laboratory	Bob Mahan Pacific Northwest National Laboratory	**DOE**
Peter W. Dean Sandia National Laboratories, Livermore	W. Frank Mason Sandia National Laboratories, Albuquerque	Thomas Ndousse U. S. Dept. of Energy, Office of Advanced Scientific Research
Walter Dykas Oak Ridge National Laboratory	Sandy Merola Lawrence Berkeley National Laboratory	Mary Ann Scott U. S. Dept. of Energy, Office of Advanced Scientific Research
Douglas E. Engert Argonne National Laboratory	James Rome Oak Ridge National Laboratory	**Universities**
Michael Fisk Los Alamos National Laboratory	Jim S. Rothfuss Lawrence Berkeley National Laboratory	Dennis Gannon Computer Science Dept., Indiana University
J. D. Fluckiger Pacific Northwest National Laboratory	Mary R. Thompson Lawrence Berkeley National Laboratory	Sara Matzner Applied Research Laboratories, The University of Texas at Austin
Barry Hess Sandia National Laboratories, Livermore	Cullen Tollbom Pacific Northwest National Laboratory	Barton Miller Computer Science Dept., University of Wisconsin - Madison
Keith R. Jackson Lawrence Berkeley National Laboratory	Steve Tuecke Argonne National Laboratory	Clifford Neuman Information Sciences Institute, University of Southern California
Kyran B. Kemper Los Alamos National Laboratory	Michael O. Vahle Sandia National Laboratories, Albuquerque	Thomas Shields CERIAS, Purdue University
Paul Krystosek CIAC, Lawrence Livermore National Laboratory	John Volmer Argonne National Laboratory	
Bob Lukens Jefferson Laboratory	Ronald Wilkins Los Alamos National Laboratory	

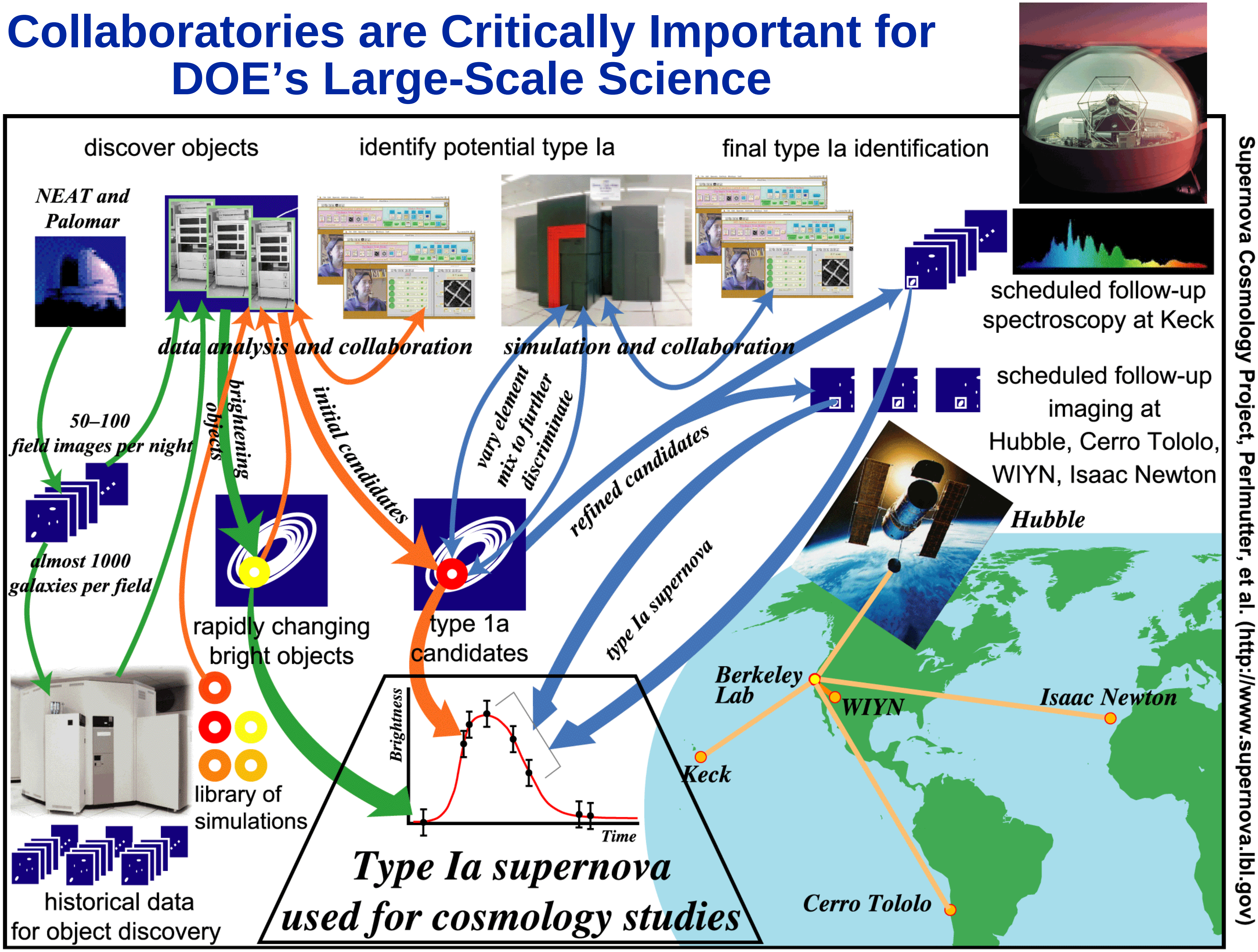
Most “Big” Science is Completely Dependent on Large Collaborations

- Sixty years ago, E. O. Lawrence pioneered the close collaboration between science and engineering that resulted in the National Labs – institutions that could address very large-scale science problems
- Today, most “big” science is dependent on world-wide collaborations that are based on
  - the free flow of data and information, and
  - easy access to remote computing, storage, and network based instruments

Collaboratories

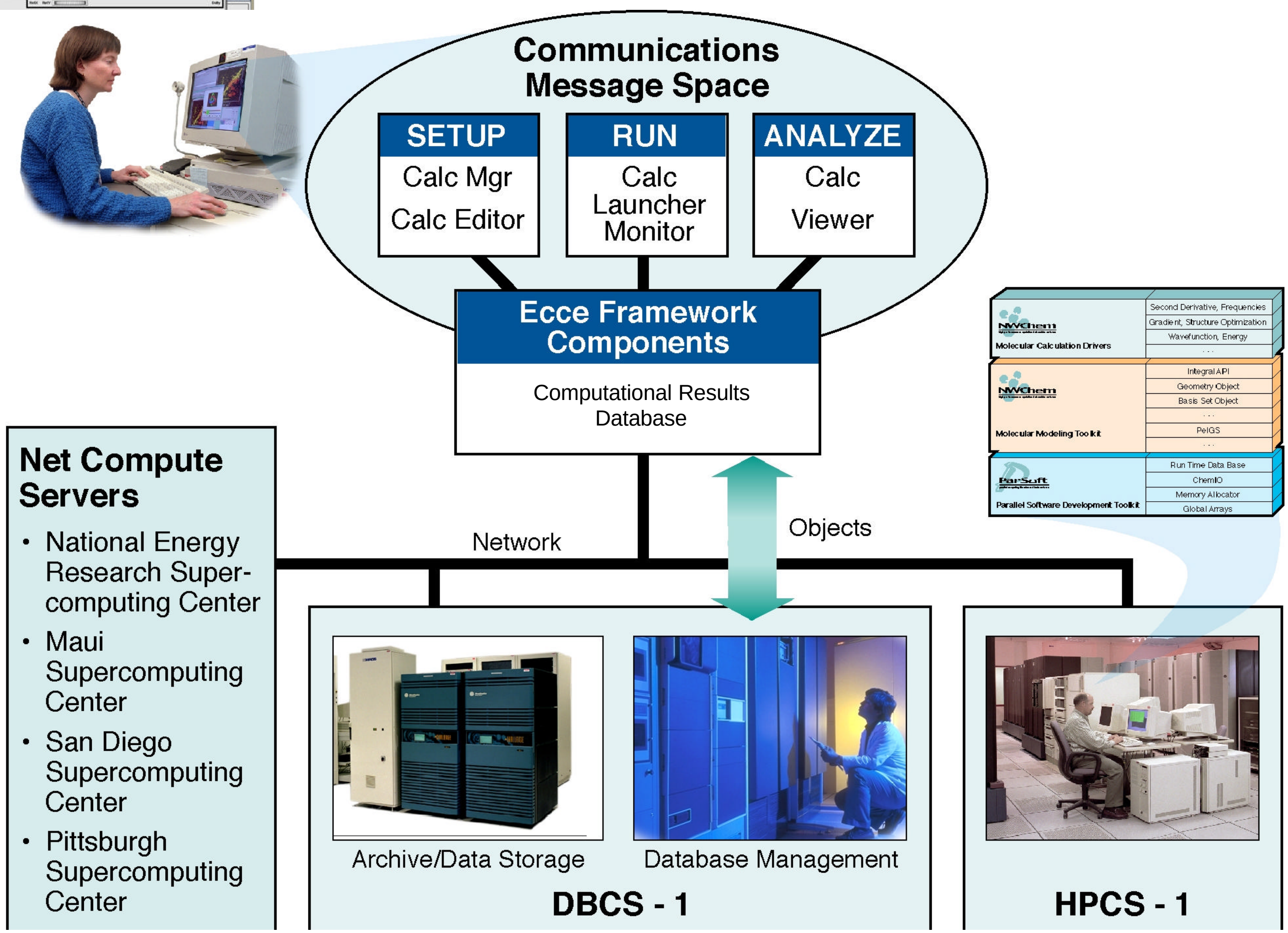
- ***Collaboratories*** are the combination of human collaborators, computer mediated services, and compute, data, and instrument resources drawn from all over the world that support the large-scale collaborations that are necessary to address the hard science problems that are at the core of DOE’s Office of Science mission
- Change is the norm in this environment, not the exception: new computing and data services are continually being developed to meet new challenges and more effectively apply computing and data analysis to solve scientific problems - rapid prototyping of digital services is how this is done

Collaboratories are Critically Important for DOE's Large-Scale Science

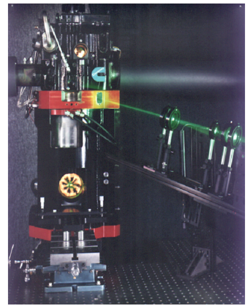


A loose knit collaboration that manages data and control for a world-wide collection of instruments.

Collaboratories and Grids Provide Access to Remote, High-End Resources



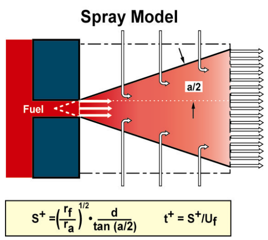
Multi-Disciplinary, Multi-Organization Collaboration is Essential



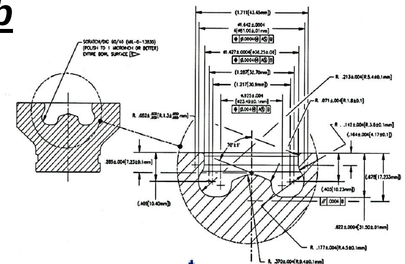
Experiment
- Laser diagnostics
- Image processing
- Scientific data management

Sandia National Lab

Physical Submodel Development
- Chemkin codes
- HCT code
- Multiphase fluid dynamics
- Combustion dynamics



Lawrence Livermore National Lab

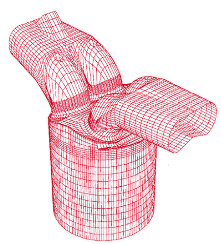


Engine Design Concepts
- Optical engines & cells
- Industry engine design
- Integrated engine tests

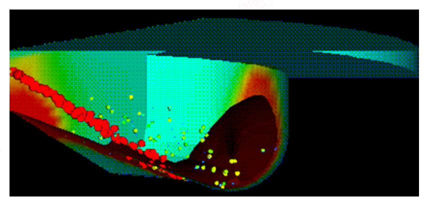
Caterpillar, Cummins Engine Company, Detroit Diesel

Diesel Combustion Research

Comprehensive Modeling
- KIVA and CHAD
- Commercial CFD codes
- MPP computing
- 3D gridding



Los Alamos National Lab



Validation Analysis
- Visualization
- Design & analysis of computer experiments
- Model data management

Univ. of Wisc.

Policy drives strong access control for some data in this collaboration

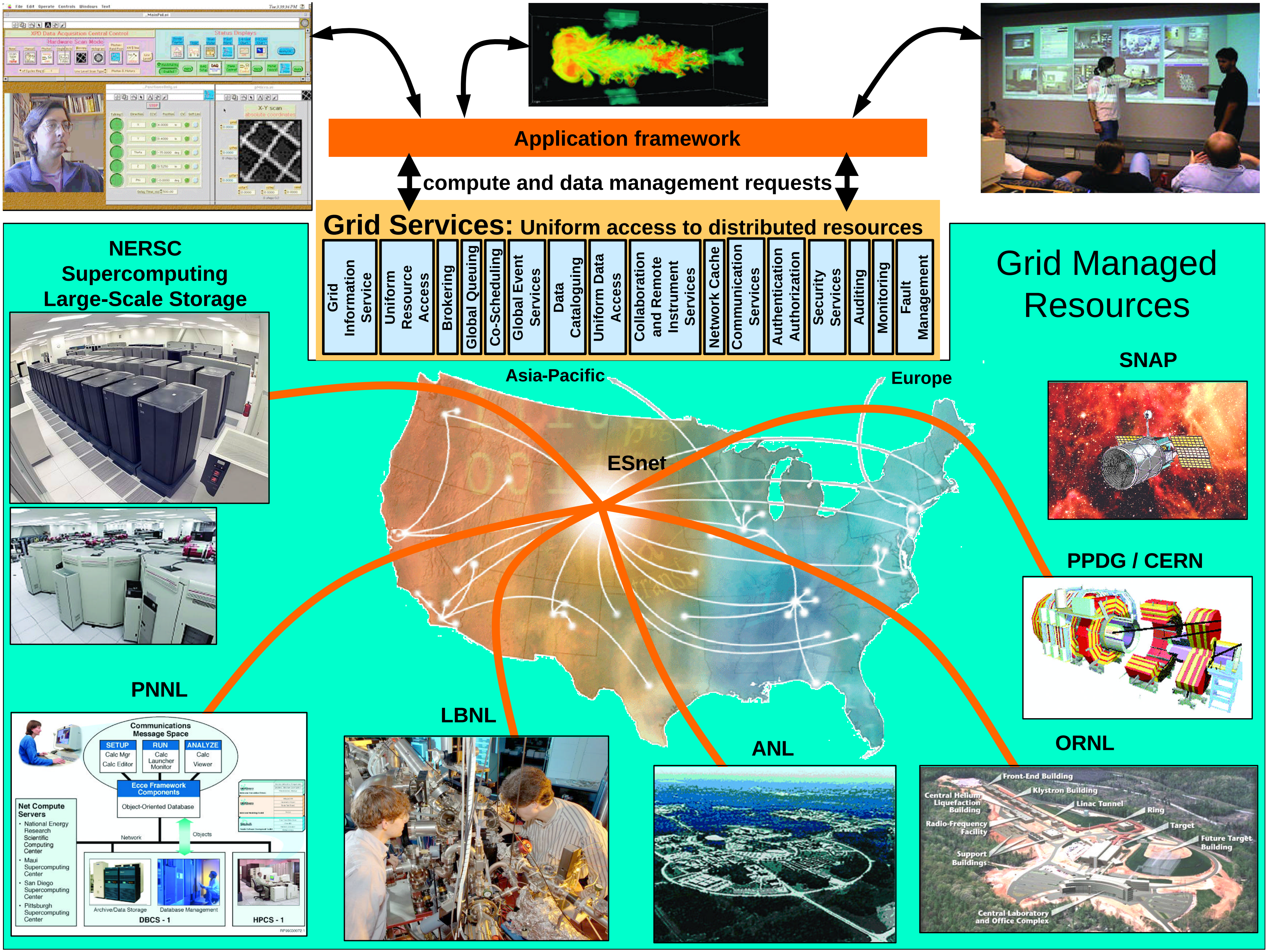
Grids:
A New Type of Infrastructure

- Grid services providing access to resources used by scientific communities
  - uniform CPU access, resource discovery, resource management, uniform data archive access, security, …

  will be the Internet Services for 21st Century science

- Deployment of this new infrastructure is underway by NASA, NSF, DOE ASCI, UK eScience Grid,
  EU Data Grid, ...

- The DOE Science Grid will provide groundwork to support unique requirements of DOE Science applications, e.g., large data, instruments, etc.



Application framework
compute and data management requests
Grid Services: Uniform access to distributed resources
Grid Information Service
Uniform Resource Access
Brokering
Global Queuing
Co-Scheduling
Global Event Services
Data Cataloguing
Uniform Data Access
Collaboration and Remote Instrument Services
Network Cache
Communication Services
Authentication Authorization
Security Services
Auditing
Monitoring
Fault Management
Grid Managed Resources
NERSC Supercomputing Large-Scale Storage
Asia-Pacific
Europe
ESnet
SNAP
PPDG / CERN
PNNL
LBNL
ANL
ORNL
Communications Message Space
SETUP
Calc Mgr
Calc Editor
RUN
Calc Launcher Monitor
ANALYZE
Calc Viewer
Ecce Framework Components
Object-Oriented Database
Net Compute Servers
National Energy Research Scientific Computing Center
Maui Supercomputing Center
San Diego Supercomputing Center
Pittsburgh Supercomputing Center
Network
Objects
Archive/Data Storage
Database Management
DBCS - 1
HPCS - 1
Front-End Building
Klystron Building
Central Helium Liquefaction Building
Linac Tunnel
Ring
Radio-Frequency Facility
Target
Future Target Building
Support Buildings
Central Laboratory and Office Complex

Motivation for Security R&D for “Open” Environments

- Collaboratories – the result of software/hardware frameworks that knit together geographically and organizationally dispersed researchers, computing systems, data, and scientific instruments – are critically important for DOE’s large-scale science
- Grids – software for uniform access to widely distributed computing, data, and instrument resources – will provide the services for building the Collaboratory frameworks that coordinate the complex application, data, instrument, and human interactions that enable large-scale science
- Security – denial of service, access control, confidentiality – is a major concern that must be addressed for viable Collaboratories, but it cannot impede the free flow of ideas and information, and access to computing resources
- A rich set of computer mediated services is critical for collaboratories: security cannot be obtained by exclusion of all but the few most common services
- DOE can make a major contribution to realizing collaboratories by defining and implementing appropriate security that protects AND allows widely distributed collaboration at the same time

Characteristics of Open Scientific Environments

- Fuzzy administrative boundaries are the norm
- International collaborations involving foreign nationals are the norm
- Variations in identity policy are the norm
- Achieving the highest possible bandwidth data flows between institutions may mean the difference between success and failure of a scientific data analysis strategy
- Scientific collaborations always consist of trusted and un-trusted systems

Security and Open R&D Environments

- Rules for high security installations cannot just be relaxed and applied to open environments - the issues are different:
  - access to computing and data is frequently much more important than confidentiality of data: protection of service is a key issue
- Experimentation with new services, new modes of communication, new mechanisms for sharing and jointly analyzing data, etc., is essential and on-going
  - scientific computing environments must be unrestricted by default, and restricted only as necessary so as to permit new services

Security and Open R&D Environments

- None the less, cybersecurity is essential in open scientific environments:
  - protection of reliable access to computing, data, and instruments is critical
    - denial-of-service attacks can be devastating - they can disrupt experiments that required months of setup time, impede a world-wide workflow of scientific data processing, etc.
  - protection against theft of service is critical
    - CPU cycles and storage space are always in short supply in the scientific community
  - access control is critical for some resources
    - delicate and expensive on-line instruments
    - some data is confidential

Security Policy and Open Environments

- Open, scientific computing environments are fragile and easily disrupted, both by hackers and by inflexible security policy and procedure
  - principles for ***protecting*** (rather than securing) the open research environment must be established and communicated to policy makers
  - collaboratory needs rarely fit well with current security policy and infrastructure
  - technical issues are raised by security policies – sometimes deliberately and sometimes as side effects

Future Computing Environment

- Ubiquitous computational and data Grids will provide the services for remote resource access and management
- Distributed problem solving environments, collaborative workbenches/frameworks, and Web portals will provide the mechanisms for accessing, expressing and managing complex scientific workflow
- New group communication services will support the larger and increasingly heterogeneous collaborations that are becoming the norm as science problems get “harder”
- Mobile access is becoming widely used and will enable scientists to maintain much closer contact with their collaborators, experiments, computer simulations, etc.

Future Threat Environment

- Mobile code, mobile workers, and wireless devices present new security challenges
- Vastly more computing power is available to hackers
- The hacker community is rapidly expanding to a worldwide scope
- Hackers are increasingly sophisticated
- Political agendas increasingly motivate hackers (e.g. in vigilante attacks on other countries)

Future Threat Environment: Increasingly Sophisticated Hackers

Considering, e.g., stacheldraht and Lion Internet Worm, it is clear that we are facing an evolution of attack scenarios:

- Autonomous: Every compromised system becomes a hacker platform for exploration, intelligence, and action, all of which are conducted autonomously
- Intelligent: Comprehensive, well though out, adaptable action plans are being conducted by autonomous systems
- Clever: Attack designers will study the MOs of Firewalls and IDSs and operate in their weaknesses, e.g. port scans conducted randomly over a very long time – say, months – so as to appear as random / uncorrelated events.
- Secretive and deceptive: Use of compressed and encrypted communication for both attack tool control and intelligence transmission

What Needs to be Done? R&D Topics

There are many issues, and therefore many security R&D topics for scientific collaboration / collaboratories

- General considerations
  - Different levels of protection for a diverse collection of resources and uses
  - Scaling solutions to 10s or 100s of institutions and 1000s of organizationally heterogeneous collaborators
  - Dynamic and static collaborations
  - Ease of use is critical
  - Cost of deployment and operation is critical
  - Collaborations have very valuable resources whose service is valuable, and must be protected and accounted
  - Some intellectual property needs to be protected

R&D Topics

- Accountability
  - Secure auditing for accounting and forensics
- Authentication
  - What's different about authenticating in a collaborative environment?
  - Multiple security domains (enclaves) - each will have different security policies & practices
  - Managing multiple authentication authorities and their trust relationships
  - Interoperability of domains using different security technology: e.g. Kerberos, PKI, SPKI, PGP

R&D Topics

- Collaborative environments need to be extensible, including untrusted or compromised resources
  - Restricted delegation: delegate minimal rights, so that untrusted/compromised resource is constrained
  - Validating untrusted environments
  - Sandboxing of processes, machines, networks: operating untrusted entities in a trusted environment

R&D Topics

- Authorization: Policy expression and checking:
  - Resources are from different labs, organizations, and countries - resource users and resource owners are not the same
  - How to define policies in different security domains so that users and resources can easily participate in collaboration?
  - Certification Authority policy interoperability: Not going to have one credential for use everywhere. How to map between CA policies?

R&D Topics

- Perimeter Protection
  - Current use of firewalls (filtering routers + application proxies) is severely detrimental to collaboration
  - Need perimeter protection without completely closing the perimeter -"adaptive, smart perimeter protection" that blocks bad guys and admits good guys
    - e.g. dynamic configuration of firewalls via user certs or proxy/delegation certs
  - How to provide high performance communication across protected perimeters
  - How to combine intrusion detection & firewall functionality to reduce overhead

R&D Topics

- Perimeter Protection (cont.)
  - How to protect UDP & IP multicast
  - Real-time intrusion detection is critical for open environments because they are / need to be less restrictive about allowing connections
  - Distributed, intelligent intrusion detection is essential because
    - collaboratories are distributed
    - denial of service attacks are distributed
  - How are perimeters defined and how do the resulting enclaves interact automatically?

R&D Topics

- Scaling Trust Environments
  - Lots of users, lots of resources - how to avoid explicit N-user to M-resource relationships
  - Can’t be prohibitively costly or burdensome
  - Recovery from compromise - compromised certs, resources
- Ease Of Use
  - Hard to use for users and/or administrators = insecure
  - The answer to security is not simply education - scientists should not have to be security experts
  - How can joining a collaboration – establishing trust and acquiring and using certs - be as easy as filling out a form

R&D Topics

- Inter-Process Communication
  - How to protect MPI, PVM, multi-media flows, group communication / multi-cast, etc.
  - How to exploit security domain (enclave) boundaries for performance and ease of configuring collaboratories
  - High performance protection - e.g., encryption algorithms that are customized to flow type
  - Group security protocols - what happens when people enter or leave collaboration group?

R&D Topics

- Grid Information Service security model
  - A critical Grid service that lets users and problem solving frameworks find out detailed information about available resources (to determine suitability for solving a particular problem)
  - System configuration information must be protected, yet at the same time available for query - how do you answer a query without revealing the underlying information (until the user is authorized)
  - How to allow broad searches without making information globally available

R&D Topics

- Analytical Models
  - How to model the effectiveness of intrusion detection, the virulence of distributed attacks
  - How to build models that give you a level of certainty that a particular observed behavior is an intrusion or denial of service attempt
  - Can models drive automatic reaction to intrusion alerts?
- Ratings/Metrics
  - Metrics need to be developed that can be used to evaluate quality of protection, scalability, policies, intrusion detection
  - Resource security rating systems would provide automated inclusion / exclusion of systems from collaborations
  - Some users may need to use a quality of protection metric for resource selection criteria
  - Can real / practical metrics be used to validate analytical models

R&D Topics

- Code safety
  - How to specify safe code behaviors and how to analyze code for unsafe behavior?
  - Given access to large amounts of computational power, are new solutions feasible?
  - What techniques can we use to control the execution of foreign code, such that when it attempts to perform an inappropriate operation, we immediately detect it?
  - What operating system facilities can we use, and what new ones need to be developed to detect all reasonable inappropriate behaviors?
  - How can we apply “binary rewriting” to transform a foreign program into one for which we can dynamically detect an unsafe action?

Cyber-Security as Science

Metrics, measurement methodology, and models are the components of a scientific approach – can this approach be applied to security?

- Metrics – measures of the ability of cyber-security tools to verify and validate systems against security objectives and requirements.
- Modeling and Analysis – mathematical techniques to establish cyber-security performance bounds and provide for the qualitative comparison of candidate cyber-security techniques and systems.
- Computational Complexity – address cyber-security issues in related disciplines such as programming languages, computer organization and operating systems, software engineering, and network protocols design
- Trust modeling – language for expressing, validating, and modeling trust in cyber-space and in large-scale scientific collaborations.
- Environment modeling – basis of detecting and responding to subtle attacks

Conclusions

- Because of DOE's science mission and associated major scientific facilities, DOE has a leadership role in building and using large-scale collaboratory environments
- The collaborations essential for large-scale science involve sharing resources across administrative and security domains, and will not happen without approaches to security that both protect and allow access
- DOE must take a leadership role in securing these environments or they will not reach their potential for fostering new and highly productive ways of doing science

Conclusions

- Major issues in need of R&D
  - Authentication across heterogeneous domains
  - Authorization: Policy expression and checking
  - Perimeter protection and ease of authorized access to resources and performance
  - Real-time and distributed intrusion detection
  - Scaling trust environments
  - Ease of use = effective security
  - Protecting Inter-process communication beyond TCP
  - Grid Information Service security model
  - Analytical models
  - Ratings/Metrics
  - Code analysis, both source and binary, for detection and modification of unsafe behavior

Conclusions

- DOE can apply its considerable experience in security and collaboratories to address and coordinate action on these problems

Acknowledgements

- This work was funded in part by the U.S. Dept. of Energy, Office of Science, Office of Advanced Scientific Computing Research, Mathematical, Information, and Computational Sciences Division (http://www.sc.doe.gov/production/octr/mics) under contract DE-AC03-76SF00098 with the University of California.